J Rehabil Med 2001; 33: 279–283

RELIABILITY AND RESPONSIVENESS OF THREE DIFFERENT PAIN ASSESSMENTS

Thomas Lundeberg,[1,2] Iréne Lund,[1] Lisbeth Dahlin,[1,2] Elsebet Borg,[3] Carina Gustafsson,[3] Lena Sandin,[3] Annika Rosén,[4] Jan Kowalski,[5] and Sven V. Eriksson[6]

From the [1]Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, [2]Department of Rehabilitation Medicine, Karolinska Hospital, Stockholm, [3]Multidisciplinary Pain Unit, Lasarettet, Ystad, [4]Department of Odontology, Division of Maxillofacial Surgery, Karolinska Insitutet, Huddinge, [5]Department of Medicine, Infection Disease, Karolinska Hospital, Stockholm, [6]Department of Medicine, Danderyd Hospital, Danderyd, Sweden

The visual analogue scale (VAS) and ordered categorical scales, i.e. numeric rating scales (NRS), are commonly used in the assessment of pain. However, these scales are bounded by fixed endpoints and thus the range of measurement is limited. The disparity in repeated assessments of perceived pain intensity with the VAS, NRS, and electrical stimulation applied as a matching stimulus was studied in 69 patients (48 women and 21 men, 19–72 years) with chronic nociceptive or neurogenic pain. Responsiveness with transcutaneous electrical nerve stimulation (TENS) using the same measurement procedures was evaluated in the same patients. Comparison of results from the three pain assessments showed that the painmatcher is at least as reliable and responsive as VAS and NRS. None of the three measurements showed evidence for systematic disagreement and had only significant random individual disagreement. They also showed evidence for responsiveness.

Key words: pain assessment, magnitude matching, VAS, NRS, treatment, TENS.

J Rehabil Med 2001; 33: 279–283

Correspondence address: Iréne Lund, Department of Physiology and Pharmacology, Karolinska Institutet, SE-171 77 Stockholm, Sweden. E-mail: irene_lund@hotmail.com

(Accepted June 14, 2001)

INTRODUCTION

An accurate assessment of the pain level experienced by a suffering patient is of great importance in making a correct diagnosis, in indicating the appropriate therapy and in studying the response to treatment. Traditionally, the visual analogue scale (VAS) and different ordered categorical scales have been used for this purpose. However, these scales are bounded, that is, they provide a limited range of measurement, confined by fixed endpoints (1). In most cases, this makes the VAS, categorical and other bounded scales sensitive to stimulus range and spacing, while subjects tend to cover the entire range of possible answers. This effect further tends to reduce the sensitivity of a scale to a pain-control intervention comparing values before and after manipulation (1).

When assessing the effect of an intervention, it is difficult to evaluate the patient's rating of the intensity of his perceived pain, as it is open to contamination from personality variables and socio-cultural factors, and the psychological and cultural influences may be as strong as the ratings of pain intensity and unpleasantness (2). It might therefore be more meaningful to introduce a painful physical stimulus against which the patient can match the level of his perceived pain. One such magnitude matching procedure was developed by Sternbach and collaborators (3) using the sub-maximum effort tourniquet technique. They concluded that the tourniquet pain ratio seems to reflect actual pain perception, without the communicative overlay of the pain estimate. They also reported that ischaemic pain measures are highly reliable and sensitive to intervention.

Ordered categorical response variables are widely used in rehabilitation (3–11). They can be recorded from a continuous range of possible values as VAS or scales with only a few possible ordered categories (4, 10, 11). In evaluating the recorded values properly it is important to note that although the record may be a numeric one, it still has no arithmetical meaning (4–12).

We present a measurement of pain, "painmatcher", which is based on electrical stimulation applied to the skin as a matching stimulus and gives a continuous ordinal individual response with no visualized predetermined lower or upper limits. The results may therefore have greater validity than those of previous methods used for measuring pain.

The main purpose of the present investigation was to evaluate the intra-individual disagreement in assessments of pain made independently with the VAS, the numeric rating scale (NRS) and painmatcher. We also investigate the ability to catch up responsiveness to changes of perceived intensity of pain by comparing values before and after treatment with transcutaneous electrical nerve stimulation (TENS).

METHODS

Study group

Sixty-nine patients (48 women and 21 men, aged 19–72 years) participated in this study. All had chronic nociceptive or neurogenic pain of varying aetiology and were enrolled in multidisciplinary rehabilitation programmes. To avoid the influence of unspecific effects associated with the initiation of treatment, all patients had earlier responded to treatment with TENS and had been receiving the treatment regularly for more than 6 months. Informed consent was given prior to

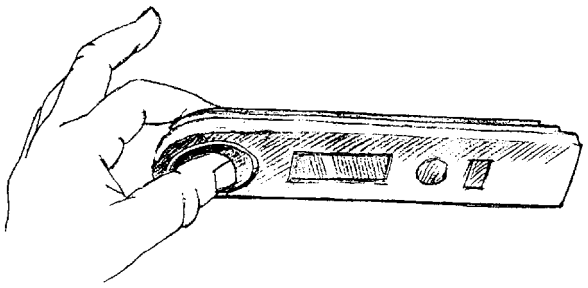

Fig. 1. Handgrip of the electrode box in the painmatcher unit.

the study, which was approved by the ethics committee of the Karolinska Hospital.

Methods

The patients were asked to rate their intensity of pain using the VAS and the NRS. They were also asked to assess the pain using a test with electrical stimulation of the skin producing magnitude matching (painmatcher). The three different assessments were presented to the patients in random order. Estimation on the VAS was recorded as a number (0–100), and the number obtained from the NRS (0–20 with 21 numerical given) is a form of quantified subjective assessment. The anchor points of both scales were at the 0-point "no pain" and at the maximum point (i.e. 100 and 20, respectively) "worst pain possible". The whole assessment procedure was then repeated. The patients were then treated with TENS for 30 min. At the end of treatment, pain levels were again assessed in order to record the immediate effects of the TENS treatment using the same procedures as before.

Magnitude matching

The electrical stimulation for "painmatcher" (the magnitude matching) was performed in the following way: The patient was instructed to place an electrode box between the first and second finger of the right hand, taking a firm grip (Fig. 1). A hand-switch was placed in the left hand. The electrical stimulation unit, delivering electrical pulses to the patient at random velocity and with increasing intensity was started by the assessor. When the sensation of pain in the right thumb and second finger corresponded in amplitude to the clinical pain, the patient could either release the fingers from the electrode box (creating an open circuit) or push the hand-switch, and a value was automatically recorded.

Magnitude matching device

The constant current electrical stimulation unit (painmatcher) is a device controlled by a microprocessor delivering rectangular pulses at a frequency of 10 Hz and amplitude of 10 mA. Intensification is achieved by successively increasing the pulse width from 0 to a possible maximum of 450 μs in increments of 7.5 μs up to a total of 60 steps. As soon as an open circuit is detected (by the release of fingers from the electrode box or pushing the hand-switch) the constantly increasing current generation halts and the value between 0 and 60 is saved in the memory. This value is directly related to the pulse width, and is displayed on an LCD screen.

Statistical analysis

The following measures have been calculated in order to evaluate reliability between first and second observations on individual data (ordinal data): systematic disagreement for the group in position (RP)

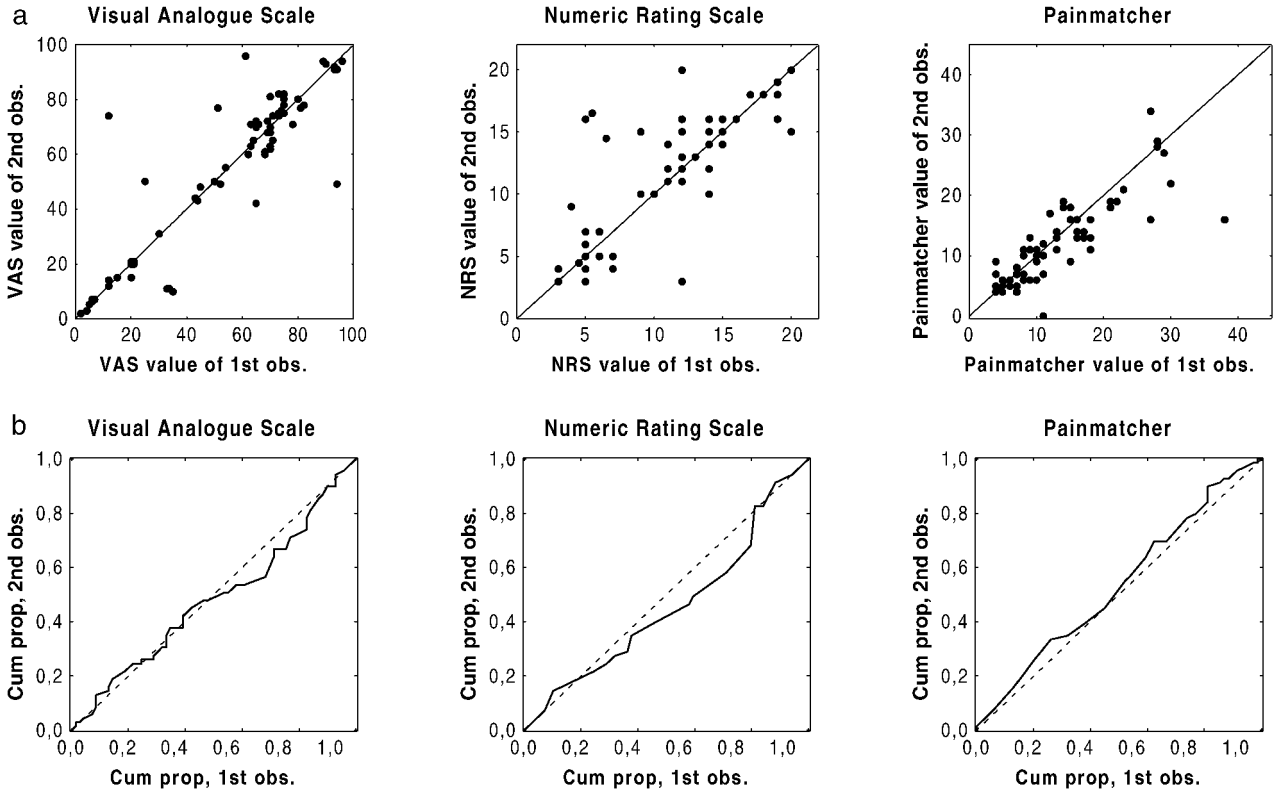

Fig. 2. Joint distribution of individual pain assessment (a) and cumulative proportion of pain assessment (b) on the visual analogue scale (VAS), numeric rating scale (NRS) and painmatcher on two occasions, first and second observation (1st obs, 2nd obs). The receiver operating characteristic (ROC) curves in (b) demonstrate the systematic disagreement in repeated assessments of pain with the three different types of measurement.

Table I. *Reliability, two repeated observations on the visual analogue scale (VAS), numeric rating scale (NRS) and painmatcher (*n = *69).*

	VAS	NRS	Painmatcher
Systematic disagreement for the group			
in position (RP) (SE)	0.032 (0.042)	0.066 (0.040)	−0.070 (0.037)
in concentration (RC) (SE)	−0.088 (0.056)	−0.049 (0.064)	0.018 (0.069)
Random individual changes (RV) (SE)	0.148 (0.057)	0.091 (0.033)	0.092 (0.024)
Augmented rank order agreement coefficient (r_a)	0.852	0.909	0.908

and in concentration (RC), level of random disagreement (RV), and the augmented rank order agreement coefficient (r_a). RP, RC, RV and r_a were also estimated as measurements for the responsiveness to the treatment with TENS (9, 10).

A high level of systematic disagreement common to the group, measured by RP and RC, indicates consistent disagreement between repeated pain assessments. The presence of additional individual changes, measured by RV, indicates heterogeneity in the group. The higher the value of r_a, between 0 and 1, the closer are the pair of augmented mean rank values to each other and the lesser is the random part of the observed disagreement (9, 10). For RV values the possible interval is from 0 to 1, and the lower the value the less the random disagreement.

Values of RP and RC close to zero indicate lack of systematic disagreement. RP is equal to 1 if there is a complete positive shift in pain assessments between two occasions. Negative RP values mean systematically lower pain recordings. Values of RC are positive if categorical distribution on the second occasion is more concentrated towards central parts of the scale compared with the distribution on the first occasion.

The receiver operating characteristic (ROC) curve was used to demonstrate the systematic disagreement in repeated assessments of pain with the three different types of measurement and for the responsiveness of TENS treatment. ROC curves were illustrated for all three measurements. Values of RP and RC close to zero indicate negligible disagreement, and the corresponding ROC curve will be close to the main diagonal (10). A concave or convex ROC curve is a sign of systematic disagreement in position of the scale and an S-shaped ROC curve is a sign of systematic disagreement of concentration of the categories.

A positive change in position will correspond to an increase in the sensation of pain from values before compared to after treatment with TENS, while a negative change in position corresponds to a decrease in the evaluation of pain. The presence of random individual changes measured by RV indicates heterogeneity in the group.

RESULTS

Fig. 2a shows the joint distribution of individual pain assessments by VAS, NRS and painmatcher on two occasions. The ROC curves demonstrating the systematic disagreement in cumulative proportion values from VAS, NRS and painmatcher are shown in Fig. 2b. Based on results shown in Table I, the estimated 95% confidence interval for systematic disagreement in position (RP) for VAS was −0.05 to 0.11, for NRS −0.01 to 0.02 and for painmatcher −0.14 to 0.01. The RP values show no evidence for statistical systematic disagreement in position, and all of them were considered small. The results from the estimation also demonstrated a lack of systematic disagreement in concentration (RC); 95% confidence intervals were for VAS −0.20 to 0.02, for NRS −0.17 to 0.08 and for the painmatcher −0.12 to 0.15. The RC value of −0.088 for VAS indicates that pain classification at the first observation was concentrated mostly in the 40–80 divisions on VAS. All three pain assessments showed statistically significant results in the individual changes in pain ($p < 0.05$), which indicates heterogeneity among the individuals. Note that the VAS demonstrated the largest value of RV. The greatest rank order agreement coefficient was 0.91 and was observed for both NRS and painmatcher.

Table II gives the responsiveness of all the scales before and after treatment with TENS, and in Fig. 3a the corresponding joint distributions are shown. There was a significant systematic change in position towards lower levels of pain and this was observed within all three measurements, whose effects are identified as a convex pattern in the ROC curves (Fig. 3b). 95% confidence intervals for RP were −0.49 to −0.26 for VAS, −0.38 to −0.16 for NRS and −0.39 to −0.14 for the painmatcher. As shown in Fig. 3a and b, no important systematic change in concentration could be seen in the ROC curve and no statistical evidence was found for the RC. In addition, there were individual changes in the pain assessments that were statistically significant for VAS, NRS and for the painmatcher. The 95% confidence intervals of the RV were 0.09 to 0.03 for VAS, 0.10 to 0.30 for NRS and 0.13 to 0.44 for the painmatcher. The random individual changes can be observed as a great variation in the joint distribution in Fig. 3a.

DISCUSSION

The results of the present study show that magnitude matching using constant current pulses in the painmatcher and ratings of

Table II. *Responsiveness, before and after treatment, with transcutaneous electrical nerve stimulation (TENS) on the visual analogue scale (VAS), numeric rating scale (NRS) and painmatcher (*n = *69).*

	VAS	NRS	Painmatcher
Systematic change for the group			
in position (RP) (SE)	−0.375 (0.061)	−0.271 (0.058)	−0.269 (0.064)
in concentration (RC) (SE)	−0.042 (0.110)	−0.066 (0.096)	0.052 (0.101)
Random individual changes (RV) (SE)	0.210 (0.060)	0.199 (0.051)	0.286 (0.079)
Augmented rank order agreement coefficient r_a	0.790	0.801	0.714

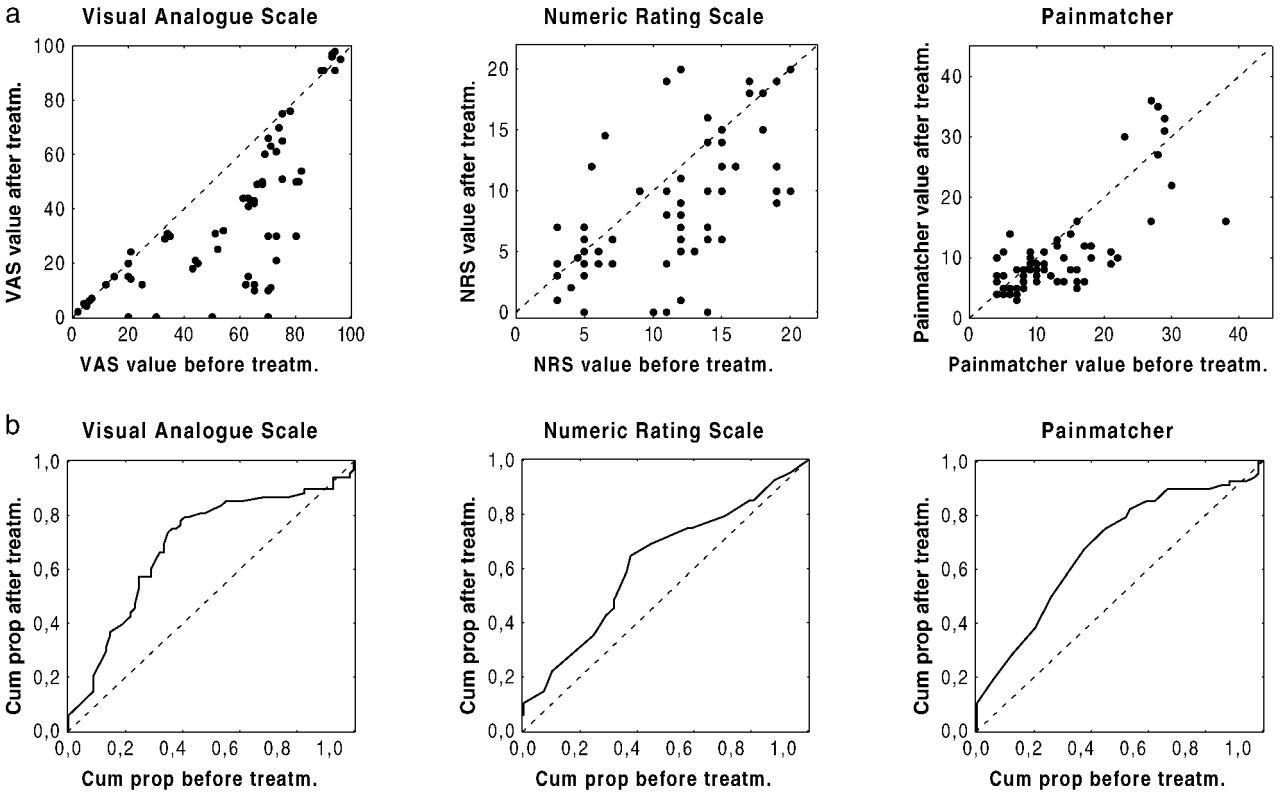

Fig. 3. Joint distribution of individual pain assessment (a) and cumulative proportion of pain assessment (b) on the visual analogue scale (VAS), numeric rating scale (NRS) and painmatcher before and after treatment with TENS. The receiver operating characteristic (ROC) curves in (b) demonstrate the responsiveness of TENS treatment in the assessments of pain with the three different types of measurement.

perceived pain using VAS and NRS all have a comparable reliability with acceptable stability in determining the perceived intensity of pain and assessing pain-controlling interventions for the group. Also, the expectations of neither the experimenter nor the patient influence the direct outcome.

The main reason for the change in pain assessment for the group compared before and after treatment with TENS was a significant systematic change in position towards lower levels of pain observed within all three measurements. This provides evidence for responsiveness in all three measurements at the group level. However, individual randomly caused changes were statistically significant, contrarily indicating a lack of responsiveness. Responsiveness measured by RP is strongly and clearly demonstrated, even if the presence of RV makes it indistinct.

The disadvantage of using measures such as repeatability coefficients and similar measures on ordinal and categorical data is that they are designed for numerical data. The use of such methods leads to serious mistakes and misinterpretation because they treat the units as "real" continuous data. In this study, an even more serious mistake would have been made if the within method variation had been measured by repeatability coefficient. The variation in differences against the mean level cannot be defined easily but it is obvious that it is not constant.

There are many ways by which to experimentally induce pain in man. However, for ethical reasons pain-inducing methods must not place the subject at risk or in substantial discomfort. Ideally, the method should not produce any tissue damage, psychological injury or any other health hazard and should have no after-effects following termination of the experiment. Furthermore, the subject must have control over immediate cessation of stimulation at any point in the investigation. Electrical stimulation is probably the most common form of experimental pain induction in man. As a constant current it is convenient and easily controllable. Electrical stimulation has previously been applied to the tooth pulp (13) and the fingers (14) in examining pain thresholds.

This new method for measuring pain, the painmatcher, may reasonably be tested in further studies as a potentially unbiased way of reporting pain and can be used by visually impaired patients. Use of the VAS possibly introduces more difficulties in older patients than in younger ones (4). In addition, two steps are involved; the estimate of pain by the patient and the clinician's measurement of the patient's line. Users must also be careful not to photocopy the VAS result, since this may alter the length of the 100 mm line, making the comparison between measurements on the photocopied scale less precise (15). This will certainly affect the outcome of a parametric approach, but should not influence the rank invariant method.

Pain intensity levels assessed on different scales have been

compared in several studies (16–18). Most have been a comparison between the VAS and NRS, (4) both of which have the limitation of remembered pain ratings and great random individual change (10). In some studies the VAS produced significantly greater discrepancies than the verbal scale. The reason might be that subjects tended to overestimate their baseline pain on the VAS, while discrepancies on the verbal scale occurred in both directions (overestimation, underestimation) when reviewed as a group. In this study, we observed a lesser value in position (systematic change in position) of the pain assessment measured by VAS, compared with both NRS and painmatcher, interpreting the RP as a measure of responsiveness. This was indeed non-significant. This result is in accordance with earlier findings, indicating overestimation on the VAS.

Another use of measuring pain is an index scale which includes measurements of pain intensity, disability and physical impairment in assessing the severity of pain (1). However, this is time-consuming. Finding a valid and reliable pain intensity scale which can be easily applied in clinical and experimental studies would be important in ordinary clinical work as well as in pain research.

Earlier studies have shown a lack in responsiveness in using VAS, demonstrated by no systematic change in position and a greater RV value (>0.50) than in this study (9, 10). Some patients find the VAS hard to use and some have difficulty understanding a numeric scale. This study shows that the painmatcher scale is at least as sensitive as the other scales and loses in precision (providing an infinite number of response categories between two extremes of pain), while NRS provides only 21 response categories compared with VAS (100 categories).

However, even though a scale may have the potential for greater discrimination, such scales are not necessarily more sensitive or valid merely by virtue of the number of categories. The large number of possible outcome values in VAS may lead to a false precision (9, 10) that will be indicated by a greater value of RV. The evaluator cannot distinguish between all (e.g. 100) categories. Most earlier comparisons have been performed between different intensity scales. Our results regarding the relation between the NRS and the VAS are supported by another recent study (7).

Because of the inherent subjectivity, pain, suffering and disability are difficult to quantify. An individual's report of pain reflects multiple contributing factors, such as cultural conditioning, expectations, social factors, state of mood and perception of control (16). Whether or not measuring the pain intensity with the painmatcher lessens the influence of these factors is not obvious from the present study. Application of the method of matching pain is, for practical reasons, possibly limited, while there are some pain conditions so severe that one would not want to provoke any further. However, for much chronic pain this arrangement should work well with patients in evaluating pain.

The ranking approach method used for evaluating data in this study was preferred. It can identify two types of disagreement, both systematic and random, compared with more traditional ones like kappa statistics. In conclusion, this study indicates that the painmatcher has the potential to be a useful and reliable tool in the measurement of pain. Values estimated on the painmatcher have proved to be at least as good as the outcome measurement with VAS and NRS with respect to random individual and systematic disagreement for responsiveness. However, this could also be a sign of overestimation of the pain relief after treatment, reflecting the intention of the patient to ameliorate. Estimations with the VAS and NRS scale are probably affected by their predetermined levels, while the painmatcher is not, since its limitations are not shown to the patient. These results indicate that pain matching is a promising technique for pain measurement. However, further studies are needed to evaluate this method in other patient groups.

REFERENCES

1. Gracely RH. Studies of pain in normal man. In: Wall PD, Melzack R, editors. Textbook of pain, 3rd edn. London: Churchill Livingstone; 1994. p. 315–335.
2. Price DD. Psychological and neural mechanisms of the affective dimension of pain. Science 2000; 288: 1769–1772.
3. Sternbach RA, Deems LM, Timmermans G, Huey LY. On the sensitivity of the tourniquet pain test. Pain 1977; 3: 105–110.
4. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. Pain 1986; 27: 117–126.
5. Carlsson AM. Assessment of chronic pain. I. Aspects of the reliability and validity of the visual analogue scale. Pain 1983; 16: 87–101.
6. McCormack HM, Horne D, Sheather S. Clinical applications of visual analogue scales: a critical review. Psychol Med 1988; 18: 1007–1019.
7. DeLoach LJ, Higgins MS, Caplan AB, Stiff JL. The visual analogue scale in the immediate postoperative period: intrasubject variability and correlation with a numeric scale. Anaest Analg 1998; 86: 102–106.
8. Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales for chronic and experimental pain. Pain 1983; 17: 45–56.
9. Svensson E. Application of a rank-invariant method to evaluate reliability of ordered categorical assessment. J Epid Biostat 1998; 4: 403–409.
10. Svensson E. Ordinal invariant measures for individual and group changes in ordered categorical data. Statistics Med 1998; 17: 2923–2936.
11. Merbitz CT, Morris J, Grip JC. Ordinal scales and foundations of misinference. Arch Phys Med Rehab 1989; 70: 308–312.
12. Thomeé R, Grimby G, Wright BD, Linacre JM. Rasch analysis of visual analog scale measurements before and after treatment of patellofemoral pain syndrome in women. Scand J Rehabil Med 1995; 27: 145–151.
13. Chapman CR, Chen AC, Bonica JJ. Effects of intrasegmental electrical acupuncture on dental pain: evaluation by threshold estimation and sensory decision theory. Pain 1977; 3: 213–227.
14. Wolf SL, Cohen BA. Sensory threshold measurements for electrical stimulation of the digits. Arch Phys Med Rehab 1977; 58: 127–132.
15. Huskinsson EC, Measurement of pain. J Reumatol 1982; 9: 768–769.
16. Thomas T, Robinson C, Champion D, McKell M, Pell M. Prediction and assessment of the severity of post-operative pain and satisfaction with management. Pain 1998; 75: 177–185.
17. Kremer E, Atkinson JH, Ignelzi RJ. Measurement of pain: patient preference does not confound pain measurement. Pain 1981; 10: 241–248.
18. Turk DC, Okifuji A. Assessment of patients reporting of pain: an integrated perspective. Lancet 1999; 353: 1784–1788.